Hopalong Casualty

Capabilities and Limitations of Visual Surveillance

Ingo Lütkebohle

Computational Perception Lab
Applied Computer Science Group
Bielefeld University

27. Dezember 2005

Visual Motion Analysis

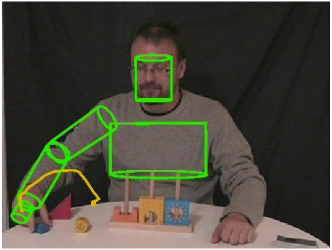
Goal: Compact description of motion.

Various levels:

- body configuration
- motion path
- "operate on block"

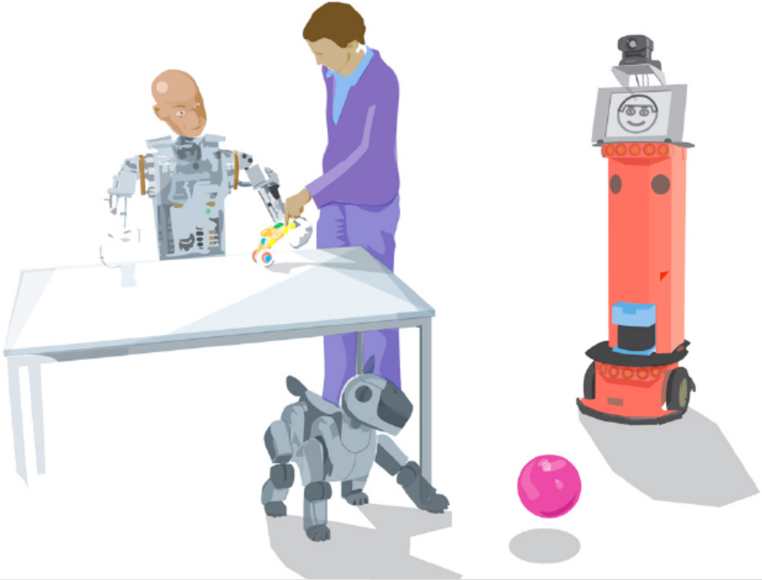
Application Areas

- Human-Computer Interaction
- Games (e.g., PS2 EyeToy)
- Motion Capture (for movies)
- *Surveillance*



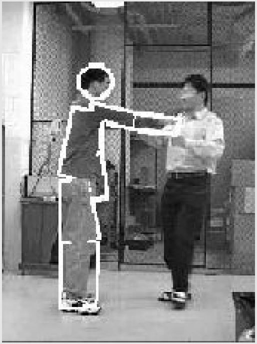
Contents of the talk

Our scenario



Why this is difficult

Ambiguity



Low Resolution

Occlusion

The *Roadrunner problem*

when you see it, it's too late already

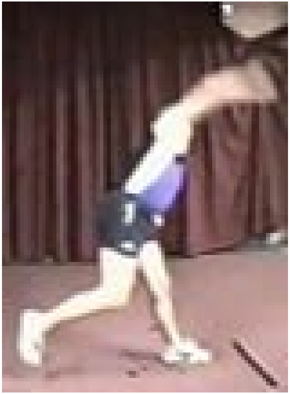
Appearance is not enough

1. Take visual experience
2. Add world knowledge
3. **Predict** activity

Human Visual Analysis

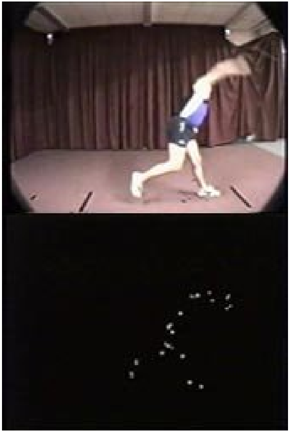
- model-based vision
- resolves visual ambiguity
- learn from visual and
- *motor* experience

Human Visual Analysis



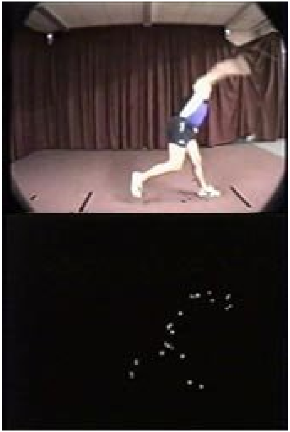
- model-based vision
- resolves visual ambiguity
- learn from visual and
- *motor* experience

Human Visual Analysis

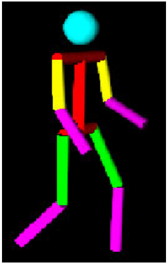


- model-based vision
- resolves visual ambiguity
- learn from visual and
- *motor* experience

Human Visual Analysis



- model-based vision
- resolves visual ambiguity
- learn from visual and
- *motor* experience



Surveillance Applications

Restricted Areas

- Little activity
- Presence detection
- Use cases:
  - Alarm trigger
  - Forensic use
- needs storage for weeks

Public Areas

- Continuous activity
- Separation, classification
- use cases
  - deterrent
  - investigative
- needs storage for days

Surveillance Specifics

Conditions

- low resolution
- low frame rate
- long stretches of nothing going on

Goals

- Categorize behaviour
- Levels
  1. regular vs. irregular
  2. run - fight - chase

Task Sketch

Computer View

- image: block of pixels (numbers)
- everything the same

Goal

- Teach a computer to detect *relevant* image parts.
- *Interpret* it

First Approach: Motion Detection

Look for *large enough* changes from one frame to the next.

Pro

- easy and fast
- gets rid of static parts

Cons

- purely intensity/color
  $\longrightarrow$ homogenous parts acquire holes
- overlaps create ambiguity

First Approach: Motion Detection

Look for *large enough* changes from one frame to the next.

Pro

- easy and fast
- gets rid of static parts

Cons

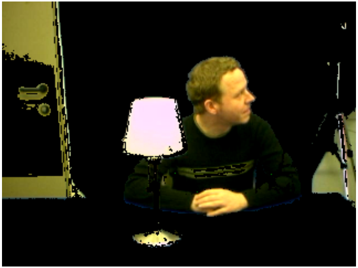
- purely intensity/color
  $\longrightarrow$ homogenous parts acquire holes
- overlaps create ambiguity

Prevent holes: Learn how background looks like

Reference Image

Input Image

Result Image

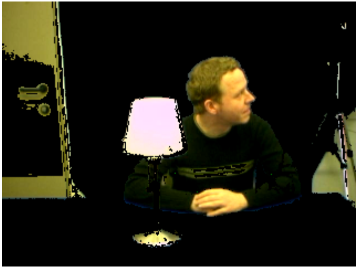


Gotcha

Prevent holes: Learn how background looks like

Reference Image

Result Image



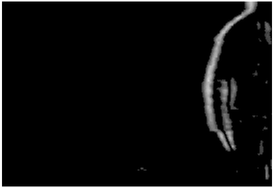
Input Image

Gotcha

Tracking to resolve ambiguities and overlap

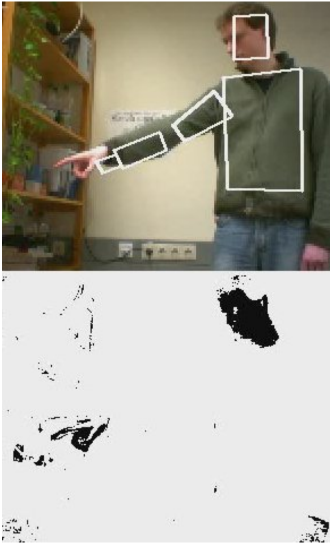
Tracking Procedure

1. First frame: Find interest points
2. Compute unique description
3. Subsequent frames: Rediscover by
    - similarity
    - proximity to expected location



Similarity: Color

- color distribution
- can focus on hands & face
- large variation
  $\longrightarrow$ silhouette as constraint
- rediscover by proximity
  $\longrightarrow$ not robust

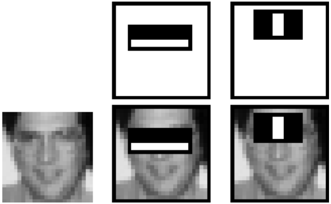


Similarity: Color

- color distribution
- can focus on hands & face
- large variation
  $\longrightarrow$ silhouette as constraint
- rediscover by proximity
  $\longrightarrow$ not robust

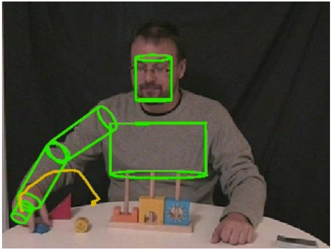
Similarity: Appearance

- “looks like” (face image)
- Look for best match
- Generalization:
  Collection of generic patches
- Very (sometimes too) specific
- Problems with rotation



Similarity: Model prediction

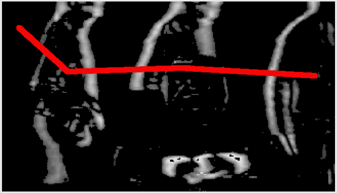
- Estimate possible positions
- Look for best match
- How to start?
- Large views only



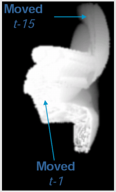
Tracking Results

Associated Postures

Trajectories



Summaries



- No intrinsic meaning
- Ambiguous

Machine Learning Approach

General Approach

1. Gather *examples* for training
2. Categorize as desired
3. Compare new images to examples
4. Assign most likely category

Challenges

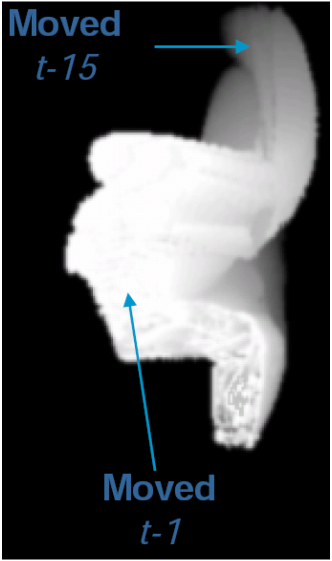
- Appearance $\neq$ function
- Duration varies
- Context matters
- What is a category anyway?

Posture

- Idea: Some postures are unique
- Find these *key* postures
- Self-occlusion problematic
- Context big part of interpretation

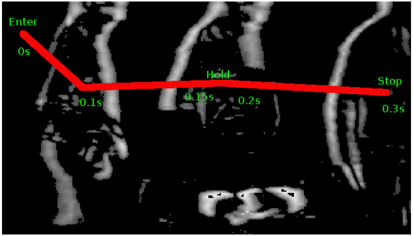
Motion History Images

- Inspired by human peripheral vision
- Compare to example images
- Only for large motions
- Requires sufficient resolution
- View-angle specific



Trajectories

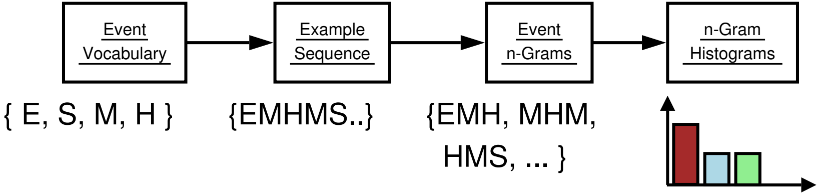
- Position (center of mass)
- Velocity, duration
- Low resolution OK
- Not much information left



Task Scripts: Recognizing abstract activities

Event Triples

- Capture context
- Fixed sample size
- Event types selected manually



Tracking Summary

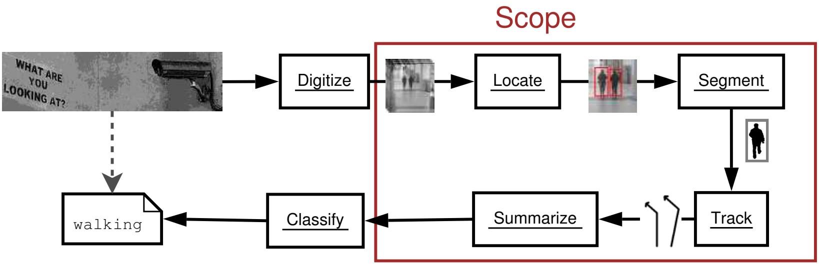
State of the Art

- Tracking associates objects over time
- Fails relatively often (even in humans)
- Robust approaches yield little information
- No clear decision between relevant and irrelevant

Results

- Hard problem for recognition
- State-of-the-art progresses fast
- Sequences not learned, yet

System Summary



For more details on camera technology, see “Hacking CCTV”, right after this talk.

Cautious note on implementations

- production software not available
  $\longrightarrow$ use research implementations, where available
- quality, robustness and speed vary
- often very particular about input data
- integration of approaches is difficult

OpenCV

Open Source Computer Vision Library

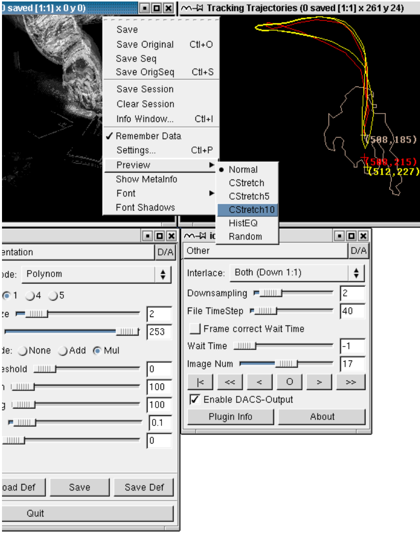
- Intel Corporation and contributors
- Comprehensive algorithm supports
- Pretty fast, can use Intel Performance Primitives (x86)
- Written in 'C', bindings for Python
- Supported on Win32 and Linux
- Main drawback: Just a library

http://www.intel.com/technology/computing/opencv/

iceWing

Open source *integration environment* for algorithms

- Basic algorithms included
- Extension via plugins, operating in a processing chain
- FireWire, V4L, AVIs, PNGs, . . .
- Plugins in 'C', C++, Python or Matlab
- Various unices and Mac OS X



http://icewing.sf.net/

Conclusion

- Indoor presence detection works
- The rest is a world full of edge cases
- Current methods are not robust enough for public areas
- Human-like results require a lot of human help
- The Roadrunner problem will be with us for a while

"I wouldn't stake my life on this technology and I wouldn't pay for it either."

Outlook: Where is it going?

Research

- Integration
- 30 pixel man, i.e. coping with bad resolution
- Interaction analysis

Congress

- Maybe a hands-on workshop? Talk to me afterwards!

Acknowledgements

Thank you for the attention!

Credits to Frank Lömker (iceWing), Joachim Schmidt (motion capture), Britta Wrede (experimental data) and Julia Lüning (22C3).